Andrew Tran

CS 1675

Homework 1 Report

Due: 1/24/19

2a/b)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Range** | [0,17] | [0,199] | [0,122] | [0,99] | [0,846] | [0,67.1] | [0.08,2.42] | [21,81] |
| **Mean** | 3.85 | 120.89 | 69.11 | 20.54 | 79.79 | 31.99 | 0.47 | 33.24 |
| **Variance** | 3.37 | 31.97 | 19.35 | 15.95 | 115.24 | 7.88 | 0.33 | 11.76 |

2c)

Class 0: 500 instances

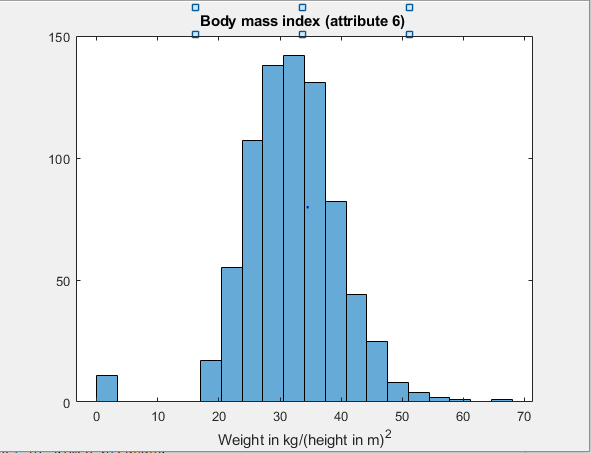
|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Mean** | 3.30 | 109.98 | 68.18 | 19.18 | 68.79 | 30.30 | 0.43 | 31.19 |
| **Variance** | 3.02 | 26.14 | 18.06 | 14.89 | 98.87 | 7.69 | 0.30 | 11.67 |

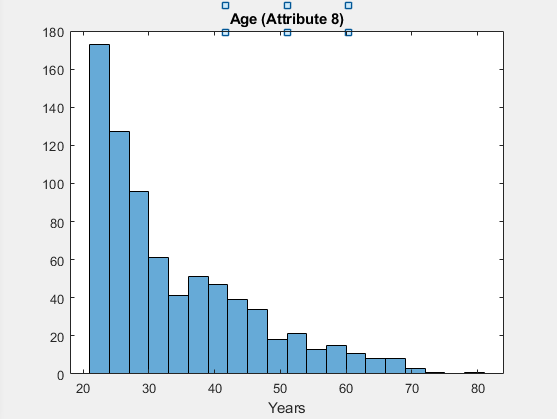
Class 1: 268 instances

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Attribute** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
| **Mean** | 4.87 | 141.26 | 70.82 | 22.16 | 100.34 | 35.14 | 0.55 | 37.07 |
| **Variance** | 3.74 | 31.94 | 21.49 | 17.68 | 138.69 | 7.26 | 0.37 | 10.97 |

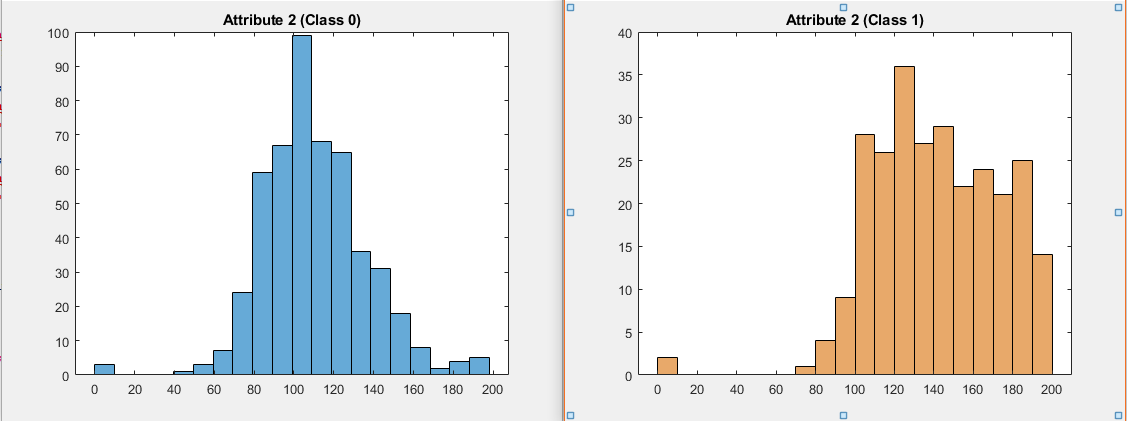
Attribute 2 (Plasma glucose concentration) seems to be the best attribute to discriminate the 2 classes. The means of this attribute have a 24.9% difference between classes which is the highest percent difference besides attribute 5. The reason attribute 5 is not as good at discriminating the 2 classes is the variances for this attribute are significantly larger than the variances for attribute 2 showing the measurement for attribute 2 is more precise.

2e) Attribute 6 (Body mass index) seems to most resemble a normal distribution.





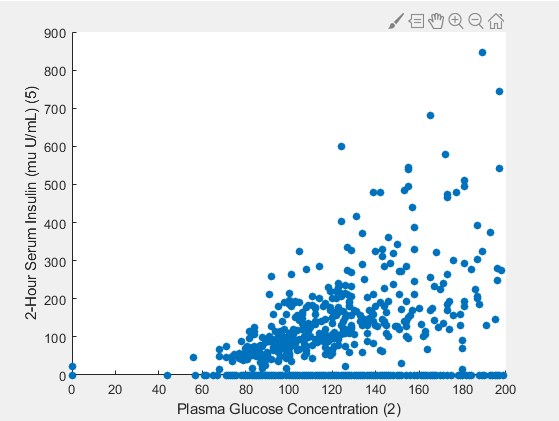
2f) Attribute 2 (Plasma glucose concentration) seems to be the best attribute the most helpful when discriminating the 2 classes:



The distribution for class 0 resembles a normal distribution while class 1’s distribution is more uniform. The distributions between the other attributes were very similar to each other between the 2 classes.

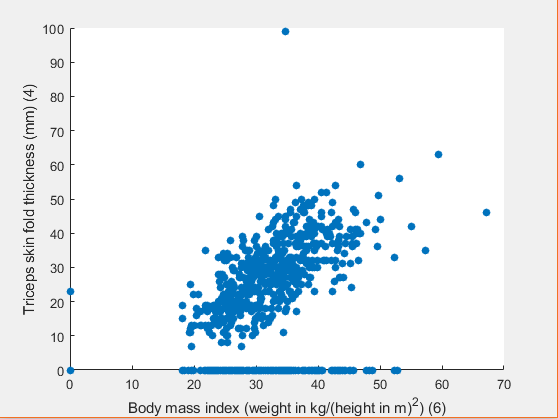
2g) If 2 attributes are independent from each other and are random, the scatter plot produced should show no obvious positive or negative trend.

An interesting pattern forms when comparing attributes 2 (plasma glucose concentration) and 5 (2-hour serum insulin):



There seems to be a positive correlation between the concentration of glucose and the amount of insulin in the bloodstream. I find this plot interesting mostly because it just makes sense. Insulin makes the body absorb glucose, so if there is a high concentration of glucose (meaning the body is not absorbing it) there should be a large amount of insulin 2 hours after taking an insulin serum (because the body is not using it to absorb the glucose).

Another non-random relationship found was between attributes 4 (triceps skin fold thickness) and 6 (BMI):



This relationship in interesting because of how compact the data points are as compared to the previous example. There is a clear positive correlation between these 2 attributes.

3a) A way to encode colors is to use a vector of 3 values (each can be either 1 or 0). This is similar to the way colors are encoded using RGB values. Since there are only 8 colors, 3 bits is enough to uniquely encode all the colors. Example:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Black | Blue | Green | Yellow | Red | Brown | Orange | White |
| Code | [0,0,0] | [0,0,1] | [0,1,0] | [0,1,1] | [1,0,0] | [1,0,1] | [1,1,0] | [1,1,1] |

3b) First 5 normalized values for attribute 3: [0.15, -0.16, -0.26, -0.16, -1.5]

3c)

|  |  |
| --- | --- |
| Original Value (attribute 3) | Discretized value (bin number) |
| 72 | 6 |
| 66 | 6 |
| 64 | 6 |
| 66 | 6 |
| 40 | 4 |

|  |  |
| --- | --- |
| 1 | 3 |
| 2 | 4 |
| 5 | 6 |

5a) A­T =

|  |  |  |
| --- | --- | --- |
| 1 | -5.5 | 1.25 |
| 0 | -0.5 | 0.25 |
| -0.67 | 4.33 | -1 |

5b) B-1 =

|  |  |  |
| --- | --- | --- |
| 15 | 7 | 14 |
| 3 | -1 | 7 |
| 3 | 6 | 10 |

5c) B+C =

|  |  |  |
| --- | --- | --- |
| -1 | -5 | 4 |
| 1 | 5 | -1 |
| 5 | 10 | 2 |

5d) B-C =

|  |  |  |
| --- | --- | --- |
| 31 | 45 | 45 |
| 53 | 59 | 75 |

5e) A\*B =

|  |  |  |
| --- | --- | --- |
| 48 | 21 | 75 |
| 15 | 0 | 30 |
| 34 | -12 | 76 |

5f) B\*C =

5g) B\*A = cannot multiply 3x3 and 2x3